XX IMEKO World Congress
Metrology for Green Growth
September 9–14, 2012, Busan, Republic of Korea

RECEDING HORIZON CONTROL FOR SELECTION OF FOCUS OF ATTENTION

M. Lauri, and R. Ritala

Department of Automation Science and Engineering, Tampere University of Technology, Tampere, Finland, P.O. Box 692, FI-33101 Tampere, Finland

Abstract: We consider the problem how a mobile robot equipped with a sensory subsystem with multiple operation modes should focus its attention to provide optimal support for finding the shortest path to a given target. The path planning problem is considered as a search task in a directed graph. The path planning is linked to the focus of attention problem by noting that any observations made of the environment ultimately reveal us information on the arc weights in the graph, i.e. the cost function. The formulation allows us to describe consistently various phenomena affecting optimal paths in the environment, for example moving obstacles that the robot may not collide with. We give a formulation for a cost function that allows trade-off between exploration of the environment and exploitation of current information by a scaling parameter.

Keywords: Focus of attention, partially observable Markov decision process, receding horizon control, dynamic programming.

1. INTRODUCTION

In the field of mobile robotics, one of the fundamental problems is planning and following optimal paths from the robot's current position to its desired goal position [1]. In a static environment, any paths required by the robot can be planned off-line before their execution. However, in a dynamically varying or a priori only partly known environment the robot must continuously update its information of the environment based on the data provided by its sensory systems and, if necessary, adjust its plans accordingly. The environment is perceived with sensors, and its state is described with a suitable map representation, e.g. a metric or topological map [2]. Sensor data may be noisy or incomplete, which introduces partial observability: we cannot determine the state of the environment from the data with certainty.

In many cases the sensory systems themselves have degrees of freedom in their operation, e.g. a camera with pan-tilt-zoom functionality may be focused to a particular area of interest. Furthermore, due to real-time requirements or limited data processing resources, not all of the available sensory data can always be processed. In order to maximize the overall performance of the system, we must select which data to process and where to focus the attention of controllable sensors. This is typically referred to as optimal control of a measurement subsystem [3] or the problem of focus of attention.

In the context of mobile robotics, the selection of focus of attention in order to maximize information of the surrounding environment and the state of the robot itself is called exploration. Information gained via exploration is, however, of no practical use unless decisions are made that exploit the information. Balancing exploration and exploitation has been widely studied in automata theory, specifically in so-called bandit problems [4], where the decision-maker gains information on the rewards of the decisions after they are made.

In many automated exploration techniques suggested for mobile robots, dynamic programming is applied to find the optimal action with respect to a cost function. The exploration problem is closely related to e.g. sensor scheduling [5], which typically results in a cost function that depends on the *information state*, i.e. a probability distribution on the state space. The information state summarises the robot's knowledge of the true state. Examples include exploration approaches in which the cost function depends e.g. on entropy [6] or related quantities [7] of the current information state, or the expected value of the information state [8].

In this paper, we consider the focus of attention problem as part of the task of planning the shortest path from an initial position to the desired goal position. Two variations of the problem are studied via simulations. The first variation applies a cost function that does not explicitly include an exploration objective, but instead implicitly optimizes the allocation of measurement resources to reach the goal position. The second variation considers a cost function dependent on the information state, where the relationship between exploration and exploitation may be determined by a scaling parameter.

2. PROBLEM FORMULATION

Let the state of the robot be $s(t) = [x(t)\ \theta(t)\ v(t)] \in S$, where t denotes the time index, x the spatial location, θ the orientation and v the velocity of the robot. The task of the robot is to plan and execute a path from an initial state $s(0)=s_{init}$ to a goal state s_{goal}.

The robot may apply on each time step a state transition action $a_s(t) \in A_s(s(t))$, which will transition the robot to a new state on the next time step according to a known deterministic function $s(t+1)=f(s(t),a_s(t))$. We assume that the robot knows with certainty its own state and the effects of its transition actions. In addition the robot has at its disposal a finite number of measurement actions $a_m(t) \in A_m(s(t))$, which may be performed concurrently with the state transition actions. Thus the total action $a(t)$ of the robot on each time step is $a(t)=[a_s(t)\ a_m(t)] \in A_s(s(t)) \times A_m(s(t))$, where $\times$ denotes the Cartesian product of the two spaces. On each time step, a cost

$$c(s(t),a(t)) = c_t(x(t))+c_s(s(t),a_s(t))+c_m(a_m(t)) \quad (1)$$

is incurred, consisting of the terrain cost of the current spatial location $c_t(x(t))$, the movement cost $c_s(s(t),a_s(t))$ and the measurement cost $c_m(a_m(t))$. From the goal state, the robot transitions into an idle state where all costs are equal to zero. For these cumulative costs, we are interested in solving

$$\min_{\{a_S(t),a_m(t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} c(s(t),a(t)). \quad (2)$$

To make a distinction between spatial locations occupied and unoccupied by obstacles, we introduce a blocking field $B(x,t)$ which takes a value 0 if location x is occupied at time t and 1 if it is unoccupied, for each location x in the set X of possible locations in the environment. For dynamically varying occupations, we may define a blocking field which evolves according to a Markov process of the form

$$p(B(x,t+1)|B(x,t)). \quad (3)$$

To account for different costs in occupied and unoccupied locations we write the robot's terrain cost as

$$c_t(x(t)) = \begin{cases} c_{terr}(x(t)) & if\, B(x,t)=0 \\ c_{coll} & if\, B(x,t)=1 \end{cases}, \quad (4)$$

where c_{coll} is the cost of collisions between robots and obstacles and $c_{terr}(x(t))$ is the terrain cost of location x. Ultimately, the observations that the robot perceives reveals information on the occupancy probability $p(B(x,t))$ of a particular location x that we focus our attention to. This may be described by writing a probabilistic observation model in the form of

$$p(o(t+1)|B(x,t+1),s(t+1),a_m(t)). \quad (5)$$

where $o(t+1) \in O$ denotes the observation.

With the above described formulation, equation (2) defines a partially observable Markov process (POMDP) control problem [9, 10], the exact optimal solution of which is known to be computationally infeasible for more than a few dozen states. Instead, we will consider here a relaxation of the full problem based on the receding horizon control principle and solving the problem on-line only for the current information state $p(B(x,t))$ of the robot. We approximate that measurements are available for a time horizon of N steps, and on subsequent time steps no new information is received. Thus we may rewrite our problem from equation (2) in terms of *expected* costs as

$$\begin{aligned} &\min_{\{a_S(t),a_m(t)\}_{t=0}^{N}} \Bigg\{ \sum_{t=0}^{N} \big[c_t(x(t))\, p(B(x(t),t)=0) \\ &+ c_{coll}(1-p(B(x(t),t)=0)) + c_s(s(t),a_s(t)) \\ &+ c_m(a_m(t))\big] \\ &+ \min_{\{a_S(t)\}_{t=N+1}^{\infty}} \sum_{t=N+1}^{\infty} \big[c_N(x(t)) + c_s(s(t),a_s(t))\big] \Bigg\} \end{aligned} \quad (6)$$

where $c_N(x(t))$ denotes the expected terrain cost based on information $p(B(x,N))$ at the final time step on which our information on the blocking field may change:

$$\begin{aligned} c_N(x(t)) &= c_t(x(t))\, p(B(x(t),N)=0) \\ &+ c_{coll}(1-p(B(x(t),N)=0)) \end{aligned}. \quad (7)$$

In equation (6), the first minimization may be solved as an on-line optimization problem for one information state, and various graph search algorithms such as Dijkstra's algorithm or A* search may be applied to find the optimal path in the static optimization problem described by second minimization term.

3. EXPECTED INFORMATION GAIN

With a modification to the cost function presented in the previous section, it is also possible to explicitly optimize the expected information gain of the various measurement actions. In this formulation, the cost function is no longer of the form presented in equation (1), but rather it depends on the occupancy information $p(B(x,t))$. Dependence on the information state results in a nonstandard POMDP formulation of the problems in equations (2) and (6), where the cost function is not anymore linear in the information state [5]. We shall now derive an expression for the expected information gain from performing action $a(t)$ in information state $p(B(x,t))$.

Let us denote by X_t the occupancy information $p(B(x,t))$ and by $H(X_t)$ its entropy. For the sake of readability, we also adopt the notation $p(B(x,t)=1) = p_{x,t}$ and drop the explicit time dependence from the state and action expressions, i.e. in the following by a, a_s, and a_m we mean $a(t)$, $a_s(t)$ and

$a_m(t)$, respectively. Furthermore, we write s and s' for $s(t)$ and $s(t+1)$, respectively.

Since the occupancies of the spatial locations are independent, the entropy of the occupancy information is

$$H(X_t) = -\sum_{x\in X} p_{x,t}\log p_{x,t} + (1-p_{x,t})\log(1-p_{x,t}) \quad (8)$$

When the robot performs action a, the occupancy information is first updated according to the dynamic model of equation (3), resulting in the occupancy information $X_{t+1}|a$. Then, the robot perceives observation o dependent on which measurement action a_m it selected and the occupancy information is updated according to the observation model of equation (5), resulting in the posterior information $X_{t+1}|a,o$.

The expected information gain $I(X_t,a)$ of action a is the difference between $H(X_{t+1}|a)$ and the expected posterior entropy $H(X_{t+1}|a,o)$ taken over O, i.e.

$$\begin{aligned} I(X_t,a) &= H(X_{t+1}\mid a) - E\left[H(X_{t+1}\mid a,o)\right] \\ &= H(X_{t+1}\mid a) \\ &\quad - \sum_{o\in O} p(o\mid X_{t+1},s',a_m)H(X_{t+1}\mid a,o) \end{aligned} \quad (9)$$

where $p(o|X_{t+1},s',a_m)$ is the prior probability for observation o, defined via the observation model of equation (5).

When applying the dynamic model of equation (3), $H(X_{t+1}|a_s)$ may either be greater than or less than $H(X_t)$. However, it holds true that $H(X_{t+1}|a) \geq \mathrm{E}[H(X_{t+1}|a,o)]$. This is due to the fact that the expression in equation (9) is equal to the mutual information between $X_{t+1}|a$ and the observations O, which by definition is always greater than or equal to zero.

Let us now define a new cost function that includes the expected information gain as

$$c(s,a,X_t) = c_t(x)+c_s(s,a_s)+c_m(a_m)-\beta\cdot I(X_t,a), \quad (10)$$

where we denote $x(t)$ by x and $\beta \geq 0$ is a scaling factor that scales the expected information gain to be comparable with the other costs. As $I(Xt,a)$ is positive, we have used a minus sign to reward for actions that result in high expected information gain.

The selection of β determines the relative importance of exploration compared to exploitation of information. An appropriate value for β may be selected experimentally or guided by considering an upper bound for $I(X_t,a)$. Let us consider the situation where we observe a single spatial location with probability p of receiving an erroneous observation of the occupancy. Let $q \in [0,1]$ be the information $X_{t+1}|a$ of the location being occupied before receiving an observation. Equation (9) is maximized when q=0.5, and thus upper bounded by

$$I(X_t,a) \leq 1 - H(O), \quad (11)$$

where we denote by $H(O)$ the entropy of the observation model:

$$H(O) = -p\log p - (1-p)\log(1-p). \quad (12)$$

Due to the independence of the occupancy probabilities, the result generalizes easily to the case where we observe K squares with independent probabilities p_i, i=1,2,…,K, of receiving erroneous observations:

$$I(X_t,a) \leq \sum_{i=1}^{K}\left(1 - H(O_i)\right). \quad (13)$$

This is an upper bound for the maximum expected information gain per time step. Equation (13) holds with equality when there is maximum uncertainty on the occupancy of the squares prior to observation.

4. SIMULATION EXPERIMENTS

We studied our approach to the focus of attention problem by solving two planning tasks, in which a robot must navigate a maze from a start location to a goal location while avoiding collisions with other entities in the environment with initially unknown location. In the first experiment, we use the problem formulation as presented in section 2, and in the second experiment we modify the cost function such that the expected information gain from measurement actions is explicitly considered, as described in section 3. Movement actions and allocation of measurement resources must be planned to complete the tasks. We shall now describe the common elements in both experiments.

We set up a maze of 15 by 15 squares, with each of the squares corresponding to one possible location in the world. A known, static blocking field was set to model permanent structures, e.g. walls of the maze. In this maze, the robot may move between adjacent spatial locations in the maze in a manner in which its orientation and velocity are not relevant. We thus make the simplifying assumption that the state of the robot consists only of its spatial location x. Four entities with random walk dynamics were set to traverse the maze to study how the robot will try to perceive them and avoid collisions. Information on their location is summarized in a probability distribution over a blocking field $B(x,t)$. The probability distribution for $B(x,t+1)$ given the probability distribution of $B(x,t)$ is

$$\begin{aligned} &p(B(x,t+1)=1\mid B(x,t)=1) = \\ &p_s\, p(B(x,t)=1) + \sum_{y\in N_x}\frac{1-p_s}{\left|N_y\right|}p(B(y,t)=1), \end{aligned} \quad (14)$$

where p_s is the probability of a random walker remaining stationary, which we set p_s=0.9, and N_x and N_y denote the sets of squares adjacent to square x and y, respectively. The notation $|\cdot|$ denotes cardinality of a set.

The measurement subsystem may be operated in eight modes, each providing an observation on the blocking state of two squares adjacent to the robot's current location. The possible observations are thus 00, 01, 10 and 11, with a probability p_1=0.1 and p_2=0.25 for erring in the primary and secondary square of focus of attention, respectively.

4.1 Experiment one

In the first experiment, we applied a cost function of the form of equation (1). For each of the squares x in the maze, a terrain cost $c_{terr}(x)$ uniformly distributed on a range from 0 to 1 was sampled. A cost of c_{goal}=-100 was set for reaching the goal. For a collision with an obstacle, a cost of c_{coll}=1500 was set. The robot may move to any of the maximum of eight squares adjacent to the square it currently occupies, with cost c_s of 0.1 times the Euclidean distance between the current square and the target square.

The cost for making any measurement was set to $c_m(a_m(t))$=0.1. When measurement costs are only considered over a finite horizon of constant length, selection between measurement actions has no effect per se on the total cost. Nevertheless, the choice of $a_m(t)$ does have an effect on the resulting occupancy information $p(B(x,t+1))$ via the observation model of equation (5), and thus the possible future costs. In other words, the various measurement possibilities have different expected utility of the information they provide with respect to solving the optimization problem of equation (6). We have chosen to include the measurement costs here to illustrate that the approach could also be applied to problems where there are measurements available with different costs.

We set an optimization horizon N=2 up to which we considered the full tree of possible combinations of movement and measurement actions and observations. When there are M_s movement actions and M_m measurement actions available with M_d possible observations on each time step, the number of nodes in this tree is $(M_s{\cdot}M_m{\cdot}M_d)^N$. As the end cost, i.e. the second minimization part of equation (6), we set the cost of the optimal path found via A* search using occupancy information $p(B(x,N))$.

The maze is shown in Figure 1. A black square denotes a square that is a priori known to be blocked and thus not traversable. An upward pointing triangle marker denotes the start location and the downward pointing triangle denotes the goal location. The path traversed by the robot when optimizing the operation of the measurement subsystem is shown with a solid red line. For reference, the optimal path found by A* search based on the initial information $p(B(x,0)=1)$=0.5 $\forall x \in X$ is indicated by a black dashed line.

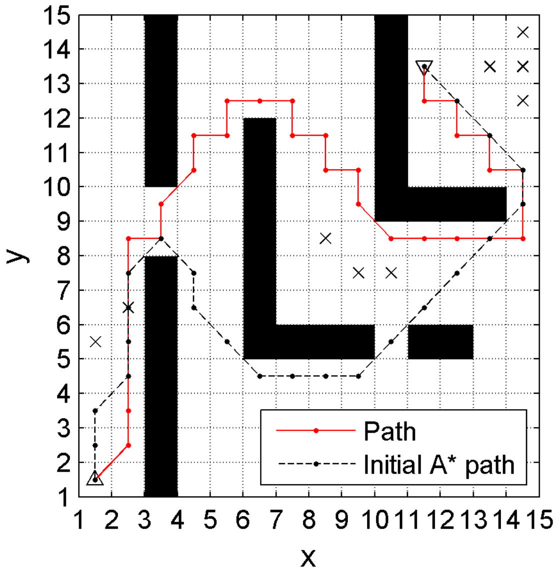

Figure 1. The maze navigated by the robot. The upward pointing triangle marks the start square and the downward pointing triangle the goal square. The solid red line denotes the actual path taken by the robot optimizing the operation of the measurement subsystem. The black line denotes the optimal path found by A* search when no measurements were available. The black crosses mark the locations where a moving obstacle was present at some time index.

To illustrate the strategy taken by the robot to detect and avoid obstacles by employing its measurement resources, we present in Table 1 the robot's location $x(t)$, primary and secondary squares $x_{o1}(t)$ and $x_{o2}(t)$ at which it focused its attention, the robots observation o consisting of two bits, and moving obstacles' locations near the robot at two distinct periods of 6 and 7 time steps. In the upper part of the table at time indices 0 through 7, the robot's preferred path leads it towards increasing values of the y-coordinate. It chooses as its primary observation direction the squares towards that direction from its current location, but at the same time tends to observe the squares towards the decreasing x-coordinate, so an evasive action may be performed in that direction if the primary direction is blocked. At time step 3 the robot observes an obstacle at square [1 5], but this does not affect its path leading towards the goal.

In the second part of Table 1 for time indices 17 through 22, we have listed the locations of two nearby obstacles $k_1(t)$ and $k_2(t)$. The robot prefers to make non-diagonal moves (e.g. from [7 11] to [8 11]) instead of diagonal moves, although they are by the definition of our movement cost function c_s less costly. This is due to the possibility to focus more attention on squares in non-diagonal directions and thereby reduce the risk of collisions there. On time step 21 the robot receives an erroneous observation 10 which leads it to perform a diagonal movement to avoid the perceived obstacle at square [10 9].

Table 1. The robot's location *x(t)*, moving obstacles' locations $k_i(t)$, the primary and secondary squares $x_{o1}(t)$ and $x_{o2}(t)$ where attention was focused and the perceived observation *o(t)* as function of time *t*.

t	*x(t)*	*k(t)*		$x_{o1}(t)$	$x_{o2}(t)$	*o(t)*
0	[1 1]	[2 6]		-	-	-
1	[2 2]	[1 5]		[2 3]	[1 3]	00
2	[2 3]	[1 5]		[2 4]	[1 4]	00
3	[2 4]	[1 5]		[2 5]	[1 5]	01
4	[2 5]	[1 5]		[2 6]	[1 6]	00
5	[2 6]	[1 5]		[2 7]	[3 7]	00
6	[2 7]	[2 6]		[3 7]	[3 8]	00
…						
t	*x(t)*	$k_1(t)$	$k_2(t)$	$x_{o1}(t)$	$x_{o2}(t)$	*o(t)*
17	[7 11]	[8 8]	[9 7]	[7 10]	[8 10]	00
18	[8 11]	[8 8]	[9 7]	[9 11]	[9 10]	00
19	[8 10]	[8 8]	[9 7]	[8 9]	[7 9]	00
20	[9 10]	[8 8]	[9 7]	[9 9]	[10 9]	00
21	[9 9]	[8 8]	[10 7]	[10 9]	[10 8]	10
22	[10 8]	[8 8]	[10 7]	[11 8]	[11 7]	00

As seen from Figure 1, a similar behaviour of preferring non-diagonal movements versus diagonal movements is present in several parts of the path. This is due to the fact that the cost of making more observations of squares that will be traversed in the near future plus the expected costs-to-goal from reachable information states *p(B(x,N))* at time horizon *N* is lower than the expected costs of making less observations and proceeding in diagonal moves, with an increased risk of collisions.

4.2 Experiment two

In the second experiment, we defined the cost function to explicitly include expected information gain, as described in section 3. Otherwise we set the same parameters for the task as in experiment one. The expected information gain per time step is bounded by equation (13) by approximately 0.72 bits. Therefore, the expected information gain may contribute at most $0.72 \cdot \beta$ to the cost function. Setting $\beta = 0$ corresponds to the setup of experiment one, as the cost function of equation (10) then equals the cost function of equation (1).

To study the effect of selecting β, we simulated the robot path for values of β from 0 to 1500, ten realizations each with the same initial configuration and goal location. Figure 2 shows the task length, i.e. the number of time steps the robot requires to reach the goal location, as function of β. The dot markers indicate the median values and the upper and lower horizontal bars indicate the minimum and maximum values among ten repeated realizations. With values of β up to 500 the task length remains low with very little variation, indicating that the robot does not consider information gathering worthwhile other than what is necessary for completing the task. Costs of movement and rewards of reaching the goal outweigh the gain of information.

When β is further increased, information gathering is considered increasingly useful when measured by the associated costs. Thus the robot tends to spend more time gathering information rather than attempting to proceed directly to the goal location. This is illustrated by the increasing median task length in Figure 2. As seen from the horizontal minimum and maximum value bars in Figure 2, the variation in the task length increases when β increases. As expected, task length has a negative correlation with the final entropy of the occupancy information at end of the task, as can be seen from Figure 3.

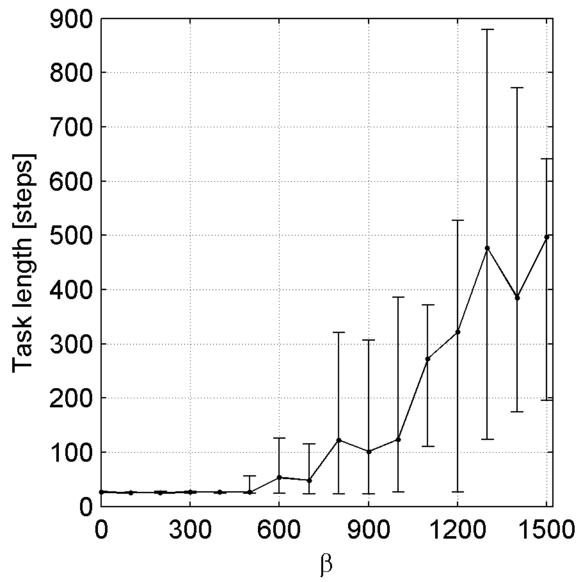

Figure 2. Task length as function of the expected information gain scaling coefficient β. The median values are indicated by the dot markers and minimum and maximum values by the horizontal bars.

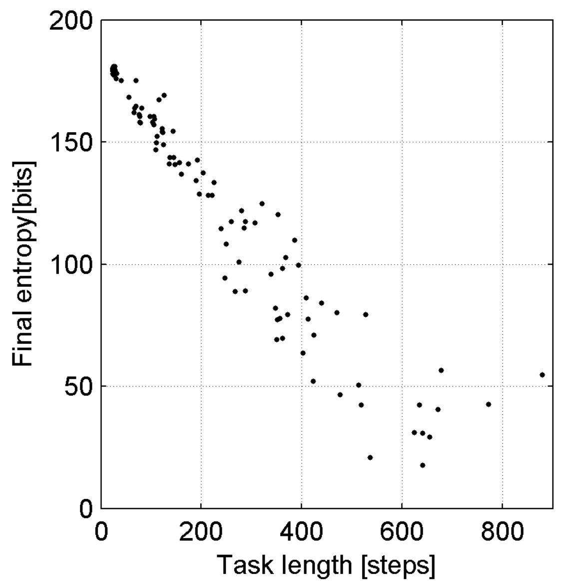

Figure 3. Entropy of the occupancy information at the end of each task versus task length.

The task length is lower bounded by the minimum number of steps required to traverse between the start location and goal location in the maze of Figure 1. There is no upper bound for the task length, and thus the robot will in some cases continue to explore the environment for an increasingly long time before moving to the goal location. However, even with values of β higher than 500, there are tasks that are completed in a low amount of time steps as with lower values of β, as shown by the minimum value bars in Figure 2. We may conclude that the scaling factor β alone does not fully determine the amount of time the robot will spend gathering information, but rather it defines in terms of costs a balance between exploratory and exploitative actions. The actual selection between exploration and exploitation behaviour depends also on the observations perceived thus far, as they affect the expected information gain from further measurement actions.

5. CONCLUSION

We have presented a formulation of the shortest path planning problem for a mobile robot as a graph search with variable arc weights. Information on the arc weights is updated by the robot's observations, and to complete the task the allocation of measurement resources must be planned in coordination with the movement actions. We also derived a formulation of the problem with a cost function that includes the expected information gain of each measurement action, weighted by a scaling parameter. Selection of the scaling parameter determines the relative importance between exploration and exploitation. The problems are solved by applying receding horizon control with a finite horizon N and estimating the expected costs from N to infinity by a graph search algorithm.

The results show that the robot selects a strategy that observes in addition to the primary direction of movement a secondary direction that allows it to avoid obstacles possibly observed in the primary direction. Varying the scaling parameter for the expected information gain was shown to bias the robot towards exploration instead of exploitation of information, although the selection between exploration and exploitation is also dependent on the observations perceived.

In the future, it could be possible to further tune the system performance by dynamically varying the scaling parameter for expected information gain as the task progresses, based e.g. on the current occupancy information state or distance to goal. This would allow more control over when the robot is behaving in exploration or exploitation mode. We also intend to demonstrate the applicability of the presented method in a real-world robot system.

6. REFERENCES

[1] S. LaValle, Planning algorithms. Cambridge University Press, 2006.

[2] B. Kuipers, J. Modayil, P. Beeson, M. MacMahon and F. Savelli, "Local metrical and global topological maps in the hybrid spatial semantic hierarchy," in Proc. of IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA, April 26 – May 1, 2004, vol. 5, pp. 4845-4851.

[3] L. Meier, J. Peschon, R. Dressler and W. G. Tuel, "Optimal control of measurement subsystems," IEEE Transactions on Automatic Control, vol. 12, no. 5, pp. 528-536, October 1967.

[4] J. Gittins, "Bandit processes and dynamic allocation indices," Journal of the Royal Statistical Society, Series B, vol. 41, no. 2, pp. 148-177, 1979.

[5] V. Krishnamurthy and D. V. Djonin, "Structured Threshold Policies for Dynamic Sensor Scheduling - A Partially Observed Markov Decision Process Approach," IEEE Transactions on Signal Processing, vol. 55, no. 10, pp. 4938-4957, Oct. 2007.

[6] A. R. Cassandra and L. P. Kaelbling, "Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation," in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Osaka, Japan, November 4-8, 1996, vol. 2, pp. 963-972.

[7] S. Thrun, "Exploration and model building in mobile robot domains," in Proc. of IEEE International Conference on Neural Networks, San Francisco, CA, USA, March 28 – April 1, 1993, pp. 175-180.

[8] R. Martinez-Cantin, N. de Freitas, E. Brochu, J. Castellanos and A. Doucet, "A Bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot," Autonomous Robots, vol. 27, no. 2, pp. 93–103, 2009.

[9] L. P. Kaelbling, M. L. Littman and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," Artificial Intelligence, vol. 101, no. 1-2, pp. 99–134, 1998.

[10] G. E. Monahan, "A survey of partially observable Markov decision processes: Theory, models, and algorithms," Management Science, vol. 28, no. 1, pp. 1–16, 1982.